

REMARKS

Claims 2-28 and 30-59 are pending after this amendment.

Claims 1 and 29 have been cancelled.

Applicants have amended claims 2-4, 6-10, 15, 20-28, 30-32, 35, 37-38, 43, 46, 48-57, and 59 in order to more particularly define the invention. No new matter has been added.

The Examiner indicated that the reference Bray, T., "Measuring the Web" which was included in an earlier-filed information disclosure statement failed to comply with the provisions of 37 CFR 1.97, 1.98 and MPEP § 609 because each publication must be identified by publisher, author (if any), title, relevant pages of the publication, date, and place of publication. Applicants are submitting here-with a supplemental information disclosure statement indicating the relevant pages 1-7 of the publication.

The Examiner objected to the specification as containing an embedded hyperlink. As suggested by the Examiner, Applicants have amended the specification to add quotation marks on each side of each hyperlink to deactivate the hyperlinks.

The Examiner rejected claims 2-3, 15, 24-25, 28, 30-31, 43, 52-53, and 56 under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which Applicants regard as their invention. Applicants have amended claims 2-3, 15, 24-25, 28, 30-31, 43, 52-53, and 56 to clarify the operation of the invention and to avoid the possibility of a continuous loop.

The Examiner rejected claims 7, 25, 28, and 35 under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which Applicants regard as their invention. The Examiner rejected dependent claims 8-12 and 36-40 for fully incorporating the deficiencies of their base claims. Applicants have amended the claims to clarify the operation of the invention with regard to adding hosts to host sets and documents to document sets.

The Examiner rejected claims 23 and 51 under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which Applicants regard as their invention. Applicants have amended claims 23 and 51 to define each element of the equation.

The Examiner rejected claims 1, 4-5, 29, 32-33, and 57 under 35 U.S.C. § 102(e) as being anticipated by Page, U.S. Patent No. 6,285,999. This rejection is respectfully traversed.

Claims 1 and 29 have been cancelled. Claim 4 has been amended so that it depends from claim 2, and thereby now incorporates all of the limitations of claim 2 as amended. Claim 5 depends from claim 4, and thereby now incorporates all of the limitations of claim 2 as amended. Claim 32 has been amended so that it depends from claim 30, and thereby now incorporates all of the limitations of claim 30 as amended. Claim 33 depends from claim 32, and thereby now incorporates all of the limitations of claim 30 as amended. Claim 57 has been amended to recite a two-level random walk, wherein upon occurrence of a random event the host selector randomly selects a host from among previously selected host. There is no hint or suggestion in Page of such a two-level random walk.

The Examiner rejected claims 13, 18-20, 41, 46-48, and 58-59 under 35 U.S.C. § 102(e) as being anticipated by Bharat et al., "A technique for measuring the relative size and overlap of public Web search engines." This rejection is respectfully traversed.

Claim 13 recites:

A computer-implemented method for measuring relative quality of a search engine index, comprising:

- a) performing a two-level random walk among documents within a document set;

- b) for each document encountered in the random walk, determining whether the document is indexed by the search engine index; and
- c) aggregating the results of b).

The claimed method thus measures relative quality of a search engine index. The method includes performing a two-level random walk to encounter a random subset of documents, and determining whether each document in the subset is indexed by the search engine. The method thus provides a mechanism for encountering documents, wherein the mechanism is independent of any search engine. Thus, the search engine being studied can be evaluated with respect to a representative sample of the entire document set rather than being compared with the results of some other search engine. Furthermore, by performing a two-level random walk, as explained in the specification, the claimed method ensures that the random walk does not unduly favor hosts having large numbers of interconnected documents at the expense of hosts having smaller numbers of interconnected documents. The two-level random walk also increases the speed of the walk in comparison with other methods. (Specification at page 15, lines 8-11).

By contrast, Bharat et al. fails to teach or disclose any such method. Bharat et al. operates by measuring overlap between search engine coverage, rather than by comparing search engine coverage with the results of a random walk. In fact, Bharat et al. states that "choosing pages uniformly at random from the entire Web is practically infeasible" (page 3), and that random walks "are not easily applicable to the Web" (page 4), thus teaching away from the very techniques claimed herein.

Rather, Bharat et al. uses the search engines themselves to generate page samples; the results of the various search engines are compared against each other in order to provide an estimate of relative overlap (pages 2, 4).

Furthermore, Bharat et al. makes no mention of any two-level random walk, as claimed herein, and would therefore fail to provide any mechanism to avoid unduly favoring hosts (domains) having large numbers of interconnected documents (web pages). As stated in the specification of the present invention, such a two-level technique provides improved, faster, and more diverse results from the random walk operation.

Claims 18-19 are dependent on claim 13, and therefore incorporate all of the limitations of claim 13. Claims 18-19 further recite additional limitations. For example, claim 18 recites a particular technique for determining whether a document is indexed by a search engine. The technique involves selecting a word from the document, performing a query on the search engine index using the word, and determining whether the document is included in the search results. This method provides a technique for determining whether a document is indexed, without having access to the search engine's databases. Claim 19 further recites that the word is selected based on rarity; this further increases the chance that, if a document is present in a search engine's database, the document will be returned as a search result when the rare word is used as the basis for a query.

Bharat et al. neither teaches nor suggests the particular techniques recited in claims 18 and 19. The cited portion of Bharat et al. (page 4, section 3) merely de-

scribes generating a random URL from a search, for purposes of generating a random sample of the contents of the search engine's database. This is an entirely different technique and rationale than that of the method of claims 18 and 19. In particular, Bharat et al. does not disclose any method of determining whether a particular document is in a search engine's database by executing a query on a word selected from the document, as claimed herein.

Furthermore, Bharat et al. fails to select a word based on its rarity, as recited in claim 19. In fact, where the Examiner stated that Bharat et al. teaches low frequency words (page 4, section 3), Bharat et al. actually discloses *exclusion* of low-frequency words from the lexicon that is being built, thus teaching away from the claimed invention.

Claim 20 as amended recites:

A computer-implemented method for measuring relative quality of a target document in a document set, comprising:

- a) performing a two-level random walk among documents within a document set; and
- b) determining a quality metric responsive to the number of times the target document is encountered in the random walk.

Claim 20 thus recites a method for measuring quality of a *document* by determining the number of times the document is encountered in a two-level random walk. The method of claim 20 thus provides an accurate assessment of overall document quality, without reference to any particular search engine. Furthermore, by performing a two-level random walk, as defined in the specification, the claimed method ensures that the random walk does not unduly favor hosts hav-

ing large numbers of interconnected documents at the expense of hosts having smaller numbers of interconnected documents. The two-level random walk also increases the speed of the walk in comparison with other methods. (Specification at page 15, lines 8-11).

Bharat et al. neither teaches nor suggests the techniques recited in claim 20. As discussed above, Bharat et al. evaluates *search engines*, rather than *documents*, by measuring overlap between search engine coverage. The cited portion of Bharat et al. (page 4, section 2.2 through page 5, section 3.1) merely describes generating a random URL from a search, for purposes of generating a random sample of the contents of the search engine's database. This is an entirely different technique and rationale than that of the method of claim 20. In particular, contrary to the Examiner's assertion, Bharat et al. does not disclose any method of determining a quality metric for a target *document* based on the number of times it is encountered in a random walk.

Furthermore, as discussed above, Bharat et al. makes no mention of any two-level random walk, as claimed herein, and would therefore fail to provide any mechanism to avoid unduly favoring hosts (domains) having large numbers of interconnected documents (web pages). As stated in the specification of the present invention, such a two-level technique provides improved, faster, and more diverse results from the random walk operation.

Claims 41 and 46-48 are computer program product claims corresponding to method claims 13 and 18-20, respectively. Claims 58 and 59 are system claims

corresponding to method claims 13 and 20, respectively. Accordingly, the above arguments apply to claims 41, 46-48, and 58-59.

Accordingly, Applicants respectfully submit that claims 13, 18-20, 41, 46-48, and 58-89 are patentable over Bharat et al.

The Examiner rejected claims 2-3, 27, 30-31, and 55 under 35 U.S.C. § 103(a) as being unpatentable over Page, U.S. Patent No. 6,285,999 in view of Singhal, U.S. Patent No. 6,370,527. This rejection is respectfully traversed.

Claim 2, which has been amended merely to more particularly define the subject matter of the invention, recites:

A computer-implemented method for randomly walking through a hypertext-linked document set comprising a plurality of documents, wherein at least a subset of the documents contain a plurality of links to other documents, each document being associated with a host, the method comprising:

- a) selecting a host;
- b) selecting at random a document associated with the host;
- c) retrieving the selected document;
- d) responsive to occurrence of a random event:
 - d.1) selecting at random a host from among the previously selected hosts;
 - d.2) selecting at random a document associated with the host; and
 - d.3) retrieving the selected document;
- e) responsive to non-occurrence of the random event:
 - e.1) selecting at random a link in the retrieved document; and
 - e.2) retrieving a document referenced by the selected link; and
- f) repeating d) and e) until a predetermined condition is met.

Claim 27, which has been amended merely to more particularly define the subject matter of the invention, recites:

A computer-implemented method for randomly walking through a hypertext-linked document set comprising a plurality of documents, wherein at least a subset of the documents contain a plurality of links to other documents, each document being associated with a host, the method comprising:

- a) selecting a host;
- b) selecting at random a document associated with the host;
- c) retrieving the selected document;
- d) responsive to occurrence of a random event:
 - d.1) selecting at random a host from among the previously selected hosts; and
 - d.2) repeating b) through e) until a predetermined condition is met; and
- e) responsive to non-occurrence of the random event:
 - e.1) selecting at random a link in the retrieved document;
 - e.2) retrieving a document referenced by the selected link; and
 - e.3) repeating d) and e) until a predetermined condition is met.

Claims 2 and 27 recite a random walk method for traversing a linked document set. The method involves selecting a host, and then selecting a document associated with the host. In using this two-level approach, the method facilitates improved performance, and furthermore reduces the degree of bias towards hosts having large numbers of interconnected documents.

The claims further recite, responsive to occurrence of a random event, selecting at random a host and selecting at random a document associated with the host. This "damping" effect causes the random walk to, at certain times, randomly select a host rather than to follow a link. By selecting the new host from among previously selected hosts, and by subsequently selecting a document associated with the host, the claimed method continues the two-level approach and avoids undesirable bias towards hosts having large numbers of interconnected documents.

Neither of the cited references, taken alone or in any combination, discloses the claimed invention. Page merely describes a technique for ranking documents in a linked database, wherein the relative importance of a page is related to the number of backlinks and the relative importance of the backlinks. Singhal merely describes a technique for submitting a search query to a plurality of search engines and compiling the results. Neither of the references contains any hint of selecting a host and then selecting a document associated with the host. In particular, neither of the references teaches randomly selecting a host, and then a document associated with the host, in response to a random event.

The Examiner correctly stated that Page does not disclose "selecting at random a host from among the previously selected hosts," as claimed herein. The Examiner cited Singhal at col. 1, lines 30-65 and col. 7, lines 21-30 as teaching selecting a search engine device for retrieval of documents. Applicants fail to see how a teaching of a search engine device has any relation whatsoever to a random selection of a host in connection with a random walk, as claimed herein. In fact, claims 2 and 27 of the present invention are not even concerned with a search engine device, but rather claim the process of the random walk without reference to a search engine. Accordingly, Applicants respectfully submit that the disclosure of Singhal is completely unrelated to the subject matter of claims 2 and 27.

Claim 3, which has been amended merely to more particularly recite the subject matter of the invention, is dependent upon amended claim 2, and incorporates all the limitations of amended claim 2.

Claims 30, 31, and 55 are computer program product claims corresponding to method claims 2, 3, and 27 respectively. Accordingly, the above arguments apply to claims 30, 31, and 55.

Accordingly, Applicants respectfully submit that claims 2-3, 27, 30-31, and 55 are patentable over the combination of Page and Singhal.

The Examiner rejected claims 6 and 34 under 35 U.S.C. § 103(a) as being unpatentable over Page, U.S. Patent No. 6,285,999 in view of Bharat et al., "A technique for measuring the relative size and overlap of public Web search engines." This rejection is respectfully traversed.

Claim 6 recites "concurrently with a) through f), performing a second two-level random walk." The claimed method thus facilitates parallel processing wherein a plurality of random walks are performed concurrently. Such an approach improves performance and efficiency of the random walk process.

Neither of the cited references, taken alone or in any combination, discloses the claimed invention. Claim 6 is dependent upon amended claim 2, and incorporates all of the limitation of amended claim 2. As discussed in connection with claim 2, Page fails to teach or suggest any technique for selecting a host and then selecting a document associated with the host, so as to implement a two-level random walk that improves performance and reduces bias. Furthermore, the Exam-

iner correctly stated that Page does not disclose “performing a second two-level random walk through the hypertext-linked document set,” as recited in claim 6.

As discussed above Bharat et al., including the cited portion at page 4, section 2.2, merely discloses techniques for measuring overlap between search engine coverage, and specifically and explicitly teaches away from the random walk techniques claimed herein. Bharat et al. states that “choosing pages uniformly at random from the entire Web is practically infeasible” (page 3), and that random walks “are not easily applicable to the Web” (page 4). In particular, Bharat et al. makes no mention whatsoever of performing a second two-level random walk concurrently with steps a) through f) recited in claim 2, as claimed herein.

Claim 34 is a computer program product claim corresponding to method claim 6. Accordingly, the above arguments apply to claim 34.

Accordingly, Applicants respectfully submit that claims 6 and 34 are patentable over the combination of Page and Bharat et al.

The Examiner rejected claims 7-12 and 35-40 under 35 U.S.C. § 103(a) as being unpatentable over Singhal, U.S. Patent No. 6,370,527 in view of Bharat et al., “A technique for measuring the relative size and overlap of public Web search engines,” and further in view of Page, U.S. Patent No. 6,285,999. This rejection is respectfully traversed.

Claim 7, which has been amended merely to more particularly recite the subject matter of the invention, recites:

A computer-implemented method for randomly walking through a hypertext-linked document set comprising a plurality of documents, wherein at least a subset of the documents contain a plurality of links to other documents, each document being associated with a host, the method comprising:

- a) initializing a host set;
- b) initializing a document set for each host in the host set;
- c) selecting at random a host from the host set;
- d) selecting at random a document from the document set of the selected host;
- e) responsive to the selected document containing at least one link:
 - e.1) selecting at random a link from the selected document;
 - e.2) selecting a document corresponding to the selected link;
 - e.3) selecting a host corresponding to the selected document;
 - e.4) adding the selected host to the host set;
 - e.5) adding the selected document to the document set of the selected host; and
 - e.6) repeating e.1) through e.5) until a first predetermined condition is met; and
- f) repeating c) through e) until a second predetermined condition is met.

Claim 9, which has been amended merely to more particularly recite the subject matter of the invention, recites:

A computer-implemented method for randomly walking through a hypertext-linked document set comprising a plurality of documents, wherein at least a subset of the documents contain a plurality of links to other documents, each document being associated with a host, the method comprising:

- a) initializing a host set;
- b) initializing a document set for each host in the host set;
- c) selecting at random a host from the host set;
- d) selecting at random a document from the document set of the selected host;
- e) responsive to non-occurrence of a random event, and further responsive to the selected document containing at least one link:
 - e.1) selecting at random a link from the selected document;
 - e.2) selecting a document corresponding to the selected link;
 - e.3) selecting a host corresponding to the selected document;
 - e.4) adding the selected host to the host set;
 - e.5) adding the selected document to the document set of the selected host; and

- e.6) repeating e.1) through e.5) until a first predetermined condition is met; and
- f) repeating c) through e) until a second predetermined condition is met.

Claims 7 and 9 recite a random walk method for traversing a linked document set. The method involves selecting a host, and then selecting a document associated with the host. If the selected document contains at least one link, a link is randomly selected from among those in the document, a document corresponding to the link is selected, and a host corresponding to the selected document is selected. In using this two-level approach, the method selects hosts and then documents, so as to facilitate improved performance, and to reduce the degree of bias towards hosts having large numbers of interconnected documents.

None of the cited references, taken alone or in any combination, discloses the claimed invention. Singhal merely describes a technique for submitting a search query to a plurality of search engines and compiling the results. Bharat et al. operates by measuring overlap between search engine coverage, rather than by comparing search engine coverage with the results of a random walk, and, as discussed above, teaches away from the random walk concept altogether. Page merely describes a technique for ranking documents in a linked database, wherein the relative importance of a page is related to the number of backlinks and the relative importance of the backlinks. None of the references contains any hint of selecting a host and then selecting a document associated with the host.

The Examiner correctly stated that neither Singhal nor Bharat discloses “selecting at random a link from the selected document, selecting a document corresponding to the selected link, and selecting a host corresponding to the selected document,” as claimed herein. The Examiner cited Page at col. 3, lines 4-16 and col. 2, lines 1-5. However, these cited portions of Page merely describe the concepts of ranking pages according to their relative importance, and using backlink information to assist in assessing relative importance. Applicants fail to see how either the page-ranking or the backlink concept can be said to anticipate the two-level selection process recited in claims 7 and 9. Neither the cited portions, nor any other portions of Page, offer any hint or suggestion of a two-level selection process, wherein a link, a document, and a host are each selected in turn, as recited in the claims.

Claims 8, 11, and 12 are dependent upon amended claim 7, and incorporate all of the limitations of amended claim 7. Claim 10 is dependent upon amended claim 9, and incorporates all of the limitations of amended claim 9.

Claims 35-40 are computer program product claims corresponding to method claims 7-12. Accordingly, the above arguments apply to claims 35-40.

Accordingly, Applicants respectfully submit that claims 7-12 and 35-40 are patentable over the combination of Singhal, Bharaet et al., and Page.

The Examiner rejected claims 14, 21-23, 26, 42, 49-51, and 54 under 35 U.S.C. § 103(a) as being unpatentable over Bharat et al., “A technique for measuring the

relative size and overlap of public Web search engines" in view of Page, U.S. Patent No. 6,285,999. This rejection is respectfully traversed.

Claim 14 is dependent upon claim 13, and therefore incorporates all of the limitations of claim 13. As discussed above, neither Bharat et al. nor Page teaches any two-level random walk as recited in claim 13 and incorporated in claim 14.

Claim 14 further recites that performing the two-level random walk comprises:

- a.1) selecting a host;
- a.2) selecting at random a document associated with the host;
- a.3) retrieving the selected document;
- a.4) selecting at random a link in the retrieved document;
- a.5) retrieving a document referenced by the selected link; and
- a.6) repeating a.4) and a.5) until a predetermined condition is met.

Claim 14 thus recites additional details concerning the selection of a host, and subsequently selecting a document associated with the host, so as to implement the two-level random walk.

The Examiner correctly stated that Bharat et al. does not disclose steps a.1) through a.5) of the claim. The Examiner cited Page, at col. 7, lines 16-21 and 38-44; col. 3, lines 4-16; and col. 2, lines 1-5 and asserted that the combination of Bharat et al. and Page renders the claim unpatentable. However, these cited portions of Page merely describe the concepts of weighting links according to distance between links, weighting a user's home page and bookmarks more highly than other pages, ranking pages according to their relative importance, and using backlink information to assist in assessing relative importance. Applicants fail to see how

any of these concepts can be said to anticipate the two-level random walk set forth in claim 14. Neither the cited portions, nor any other portions of Page, offer any hint or suggestion of a two-level random walk as recited.

Claim 21, which has been amended merely to more particularly set forth the subject matter of the invention, recites:

A computer-implemented method for measuring relative quality of a target document in a document set comprising a plurality of documents, wherein at least a subset of the documents contain a plurality of links to other documents, the method comprising:

- a) performing a two-level random walk among documents within a document set; and
- b) determining a quality metric responsive to the number of documents encountered during the two-level random walk that link to the target document.

As discussed above, the two-level random walk provides a more efficient and accurate assessment of relative quality of the target document. By employing a two-level random walk, the claimed invention is able to more quickly traverse documents, and is further able to reduce the potential for bias in favor of hosts having large numbers of interconnected documents.

As discussed above, neither Bharat et al. nor Page teaches any two-level random walk as recited in claim 21. In particular, the cited portion of Bharat et al. at page 4, section 2.2, merely discloses techniques for measuring overlap between search engine coverage, and specifically and explicitly teaches away from the random walk techniques claimed herein. Bharat et al. states that "choosing pages uniformly at random from the entire Web is practically infeasible" (page 3), and that random walks "are not easily applicable to the Web" (page 4). Bharat et al.

makes no mention whatsoever of performing a two-level random walk as claimed herein.

Claims 22, 23, and 26, which have been amended merely to more particularly recite the subject matter of the invention, are dependent upon amended claim 21, and incorporate all the limitations of amended claim 21.

Claims 42, 49-51, and 54 are computer program product claims corresponding to method claims 14, 21-23, and 26, respectively. Accordingly, the above arguments apply to claims 42, 49-51, and 54.

The Examiner rejected claims 15-17, 24-25, 28, 43-45, 52-53, and 56 under 35 U.S.C. § 103(a) as being unpatentable over Bharat et al. and Page, and further in view of Singhal, U.S. Patent No. 6,370,527. This rejection is respectfully traversed.

Claim 15, which has been amended merely to more particularly set forth the subject matter of the invention, recites:

A computer-implemented method for measuring relative quality of a search engine index, comprising:

- a) performing a two-level random walk among documents within a document set, by:
 - a.1) selecting a host;
 - a.2) selecting at random a document associated with the host;
 - a.3) retrieving the selected document;
 - a.3.1) responsive to occurrence of a random event:
 - a.3.1.1) selecting at random a host from among the previously selected hosts;
 - a.3.1.2) selecting at random a document associated with the host; and
 - a.3.1.3) retrieving the selected document;
 - a.3.2) responsive to non-occurrence of the random event:
 - a.4) selecting at random a link in the retrieved document; and
 - a.5) retrieving a document referenced by the selected link; and

- a.6) repeating a.3.1) through a.5) until a predetermined condition is met;
- b) for each document encountered in the random walk, determining whether the document is indexed by the search engine index; and
- c) aggregating the results of b).

The claimed method thus measures relative quality of a search engine index. The method includes performing a two-level random walk to encounter a random subset of documents, and determining whether each document in the subset is indexed by the search engine. The method thus provides a mechanism for encountering documents, wherein the mechanism is independent of any search engine. Thus, the search engine being studied can be evaluated with respect to a representative sample of the entire document set rather than being compared with the results of some other search engine. Furthermore, by performing a two-level random walk, as defined in the specification, the claimed method ensures that the random walk does not unduly favor hosts having large numbers of interconnected documents at the expense of hosts having smaller numbers of interconnected documents. The two-level random walk also increases the speed of the walk in comparison with other methods. (Specification at page 15, lines 8-11).

The claim further recites, responsive to occurrence of a random event, selecting at random a host and selecting at random a document associated with the host. This "damping" effect causes the random walk to, at certain times, randomly select a host rather than to follow a link. By selecting the new host from among previously selected hosts, and by subsequently selecting a document associated with the host, the claimed method continues the two-level approach and

avoids undesirable bias towards hosts having large numbers of interconnected documents.

As discussed above, none of the cited references, taken alone or in any combination, discloses the claimed invention. Bharat et al. merely discloses techniques for measuring overlap between search engine coverage, and specifically and explicitly teaches away from the random walk techniques claimed herein. Page merely describes a technique for ranking documents in a linked database, wherein the relative importance of a page is related to the number of backlinks and the relative importance of the backlinks. Singhal merely describes a technique for submitting a search query to a plurality of search engines and compiling the results. None of the references contains any hint of selecting a host and then selecting a document associated with the host. In particular, none of the references teaches randomly selecting a host, and then a document associated with the host, in response to a random event.

The Examiner correctly stated that Bharat et al. and Page do not disclose “selecting at random a host from among the previously selected hosts.” The Examiner cited Singhal at col. 1, lines 30-65 and col. 7, lines 21-30 for teaching selecting a search engine device. Applicants respectfully point out that the selection of a search engine device, as disclosed in Singhal for the purpose of aggregating search results from a plurality of search engines, is entirely different from selection of a host as part of a two-level random walk, as claimed herein. In fact, the two-level random walk, including the operation of selecting a host as claimed herein, is en-

tirely independent of selection of a search engine, and thus the disclosure of Singhal is completely unrelated to the claimed limitation.

Claim 16 is dependent upon claim 13, and incorporates all of the limitations of claim 13. As discussed above, none of the cited references teaches any two-level random walk as recited in claim 13 and incorporated in claim 16.

Claim 16 further recites that performing the two-level random walk comprises:

- a.1) initializing a host set;
- a.2) initializing a document set for each host in the host set;
- a.3) selecting at random a host from the host set;
- a.4) selecting at random a document from the document set of the selected host;
- a.5) adding the selected host to the host set;
- a.6) adding the selected document to the document set of the selected host;
- a.7) responsive to the selected document containing at least one link:
 - a.7.1) selecting at random a link from the selected document;
 - a.7.2) selecting a document corresponding to the selected link;
 - a.7.3) selecting a host corresponding to the selected document;
 - a.7.4) repeating a.5) through a.8) until a predetermined condition is met; and
- a.8) responsive to the selected document not containing at least one link, repeating a.3) through a.8) until a predetermined condition is met.

Claim 16 thus recites additional details concerning the selection of a host, and subsequently selecting a document associated with the host, so as to implement the two-level random walk.

The Examiner correctly stated that Bharat et al. does not disclose “selecting at random a link from the selected document; selecting a document corresponding to the selected link; and selecting a host corresponding to the selected document.”

The Examiner cited Page, at col. 7, lines 38-44; col. 3, lines 4-16; and col. 2, lines 1-5 and asserted that the combination of Singhal, Bharat et al., and Page renders the

claim unpatentable. However, these cited portions of Page merely describe the concepts of weighting links according by weighting a user's home page and bookmarks more highly than other pages, ranking pages according to their relative importance, and using backlink information to assist in assessing relative importance. Applicants fail to see how any of these concepts can be said to anticipate the two-level random walk set forth in claim 14. Neither the cited portions, nor any other portions of Page, offer any hint or suggestion of a two-level random walk as recited.

Claim 17 is dependent upon claim 16, and incorporates all of the limitations of claim 16.

Claim 24, which has been amended merely to more particularly set forth the subject matter of the invention, recites:

A computer-implemented method for measuring relative quality of a target document in a document set comprising a plurality of documents, wherein at least a subset of the documents contain a plurality of links to other documents, wherein each document is associated with a host, the method comprising:

- a) performing a two-level random walk among documents within a document set, by:
 - a.1) selecting a host;
 - a.2) selecting at random a document associated with the host;
 - a.3) retrieving the selected document;
 - a.4) responsive to occurrence of a random event:
 - a.4.1) selecting at random a host from among the previously selected hosts;
 - a.4.2) selecting at random a document associated with the host; and
 - a.4.3) retrieving the selected document;
 - a.5) responsive to non-occurrence of the random event:
 - a.5.1) selecting at random a link in the retrieved document; and
 - a.5.2) retrieving a document referenced by the selected link; and

- a.6) repeating a.4) to a.5) until a predetermined condition is met; and
- b) determining a quality metric responsive to the number of documents encountered during the two-level random walk that link to the target document.

Claim 24 thus recites details concerning the selection of a host, and subsequently selecting a document associated with the host, so as to implement a two-level random walk. As discussed above, the two-level random walk provides a more efficient and accurate assessment of relative quality of the target document. By employing a two-level random walk, the claimed invention is able to more quickly traverse documents, and is further able to reduce the potential for bias in favor of hosts having large numbers of interconnected documents.

As discussed above, none of the cited references teaches any two-level random walk as recited in claim 24.

The Examiner correctly stated that Bharat et al. does not disclose “selecting a host; selecting at random a document associated with the host; retrieving the selected document; selecting at random a link in the retrieved document; and retrieving a document referenced by the selected link.” The Examiner correctly stated that Bharat et al. does not disclose “selecting at random a host from among the previously selected hosts.” The Examiner cited Page, at col. 7, lines 16-21 and 38-44; col. 3, lines 4-16; and col. 2, lines 1-5, and further cited Singhal at col. 1, lines 30-65 and col. 7, lines 21-30, and asserted that the combination of Singhal, Bharat et al., and Page renders the claim unpatentable. However, the cited portions of Page merely describe the concepts of weighting links according by weighting a

user's home page and bookmarks more highly than other pages, ranking pages according to their relative importance, and using backlink information to assist in assessing relative importance. The cited portions of Singhal merely describe merging results of multiple search engines, and selecting among search engines. Applicants fail to see how any of these concepts can be said to anticipate the two-level random walk set forth in claim 24. Neither the cited portions, nor any other portions of Page or Singhal, offer any hint or suggestion of a two-level random walk as recited.

Claim 25, which has been amended merely to more particularly set forth the subject matter of the invention, recites:

A computer-implemented method for measuring relative quality of a target document in a document set comprising a plurality of documents, wherein at least a subset of the documents contain a plurality of links to other documents, wherein each document is associated with a host, the method comprising:

- a) performing a two-level random walk among documents within a document set, by:
 - a.1) initializing a host set;
 - a.2) initializing a document set for each host in the host set;
 - a.3) selecting at random a host from the host set;
 - a.4) responsive to occurrence of a random event:
 - a.4.1) selecting at random a host from among the previously selected hosts;
 - a.5) responsive to non-occurrence of the random event:
 - a.5.1) selecting at random a document from the document set of the selected host; and
 - a.5.2) responsive to the selected document containing at least one link:
 - a.5.2.1) selecting at random a link from the selected document;
 - a.5.2.2) selecting a document corresponding to the selected link;
 - a.5.2.3) selecting a host corresponding to the selected document; and
 - a.5.2.4) adding the selected host to the host set;

- a.5.2.5) adding the selected document to the document set of the selected host;
- a.5.2.6) repeating a.5.2.1) through a.5.2.5) until a first predetermined condition is met; and
- a.6) repeating a.3) through a.5) until a second predetermined condition is met; and
- b) determining a quality metric responsive to the number of documents encountered during the two-level random walk that link to the target document.

Claim 28, which has been amended merely to more particularly set forth the subject matter of the invention, recites:

A computer-implemented method for measuring relative quality of a target document in a document set comprising a plurality of documents, wherein at least a subset of the documents contain a plurality of links to other documents, the method comprising:

- a) performing a two-level random walk among documents within a document set, by:
 - a.1) initializing a host set;
 - a.2) initializing a document set for each host in the host set;
 - a.3) selecting at random a host from the host set;
 - a.4) responsive to occurrence of a random event:
 - a.4.1) selecting at random a host from among the previously selected hosts;
 - a.5) responsive to non-occurrence of the random event:
 - a.5.1) selecting at random a document from the document set of the selected host; and
 - a.5.2) responsive to the selected document containing at least one link:
 - a.5.2.1) selecting at random a link from the selected document;
 - a.5.2.2) selecting a document corresponding to the selected link;
 - a.5.2.3) selecting a host corresponding to the selected document; and
 - a.5.2.4) adding the selected host to the host set;
 - a.5.2.5) adding the selected document to the document set of the selected host;
 - a.5.2.6) repeating a.5.2.1) through a.5.2.5) until a first predetermined condition is met; and
 - a.6) repeating a.3) through a.5) until a second predetermined condition is met; and
 - b) determining a quality metric responsive to the number of documents encountered during the two-level random walk that link to the target document;
 - c) determining a quality metric for at least one additional target document; and

- d) ranking the quality metric of the first document with respect to the quality metric of the additional target document.

Claims 25 and 28 recite a random walk method for traversing a linked document set. The method involves selecting a host, and then selecting a document associated with the host. If the selected document contains at least one link, a link is randomly selected from among those in the document, a document corresponding to the link is selected, and a host corresponding to the selected document is selected. In using this two-level approach, the method selects hosts and then documents, so as to facilitate improved performance, and to reduce the degree of bias towards hosts having large numbers of interconnected documents.

None of the cited references, taken alone or in any combination, discloses the claimed invention. Singhal merely describes a technique for submitting a search query to a plurality of search engines and compiling the results. Bharat et al. operates by measuring overlap between search engine coverage, rather than by comparing search engine coverage with the results of a random walk, and, as discussed above, teaches away from the random walk concept altogether. Page merely describes a technique for ranking documents in a linked database, wherein the relative importance of a page is related to the number of backlinks and the relative importance of the backlinks. None of the references contains any hint of selecting a host and then selecting a document associated with the host.

The Examiner correctly stated that Bharat et al. does not disclose “selecting at random a link from the selected document; selecting a document corresponding

to the selected link; and selecting a host corresponding to the selected document. The Examiner cited Page, at col. 7, lines 16-21 and 38-44; col. 3, lines 4-16; and col. 2, lines 1-5, and asserted that the combination of Singhal, Bharat et al., and Page renders the claim unpatentable. However, the cited portions of Page merely describe the concepts of weighting links according by weighting a user's home page and bookmarks more highly than other pages, ranking pages according to their relative importance, and using backlink information to assist in assessing relative importance. Applicants fail to see how any of these concepts can be said to anticipate the two-level random walk set forth in claims 25 and 28. Neither the cited portions, nor any other portions of Page, offer any hint or suggestion of a two-level random walk as recited.

Claims 43-45, 52-53, and 56 are computer program product claims corresponding to method claims 15-17, 24-25, and 28, respectively. Accordingly, the above arguments apply to claims 43-45, 52-53, and 56.

Applicants have further amended the claims merely to more particularly define the subject matter of the invention. None of the amendments herein were necessitated by the Examiner's rejections. No new matter has been added.

Accordingly, Applicants respectfully submit that claims 2-28 and 30-59 are patentably distinct over the references cited.

On the basis of the above amendments, consideration of this application
and the early allowance of all claims herein are requested.

Favorable action is solicited.

Respectfully submitted,
Monika R. Henzinger and
Michael D. Mitzenmacher

Dated: March 17, 2003

By: 
Amir H. Raubvogel
Registration #37,070
Fenwick & West LLP
801 California Street
Mountain View, CA 94041
Phone: (650) 335-7276
Fax: (650) 938-5200